Multivariate Approach for Quantification of Morphologic and Functional Damage in Glaucoma

Peter Martus,¹ Anselm Jüinemann,² Martin Wisse,² Wido M. Budde,² Folkert Horn,² Mattbias Korth,² and Jost B. Jonas²

Purpose. To determine the usefulness of confirmatory factor analysis in examination of morphometric, electrophysiological, and psychophysical quantitative methods that measure the extent of global glaucomatous damage without referring to a preselected gold standard.

Methods. In a cross-sectional clinical study, 406 eyes of 203 glaucoma patients and 200 eyes of 100 normal control subjects 18 to 70 years old underwent optic disc morphometry, automated perimetry, measurement of temporal contrast sensitivity by a full-field flicker test, blue-on-yellow visually evoked potential (VEP), and black-and-white pattern-reversal electroretinogram (ERG). Diagnosis of glaucoma was based on a qualitative classification of the optic nerve head and retinal nerve fiber layer independent of intraocular pressure and visual field. Confirmatory factor analysis was performed in the patient group as a whole and in a subgroup showing moderate to advanced glaucomatous optic nerve head damage.

Results. The confirmatory factor analysis models explained the data satisfactorily (P > 0.18, all patients; P > 0.34, subgroup). Global glaucomatous damage was quantified best by the mean defect of automated perimetry (r = 0.81; r = 0.87), followed by the area of the neuroretinal rim (r = 0.64; r = 0.73), the full-field flicker test (r = 0.59; r = 0.65), the pattern-reversal ERG amplitude (r = 0.54; r = 0.55), and the VEP peak time (r = 0.55; r = 0.54).

Conclusions. Confirmatory factor analysis allows quantification of the validity of established and new procedures that measure global glaucomatous damage using cross-sectional data. The results are not dependent on the preselection of a specific gold standard. Psychophysical testing and morphometry quantified glaucomatous damage best, compared with electrophysiological procedures. (Invest Ophthalmol Vis Sci. 2000;41:1099-1110)

A multitude of measures has been developed to determine the damage of visual function caused by glaucomatous diseases. The usefulness of these measures may be evaluated by various criteria: early detection of glaucoma, quantification of the glaucomatous damage, early detection and quantification of the progress of the disease, and prognosis of the disease. The present study was focused on the second of these criteria: To determine how different measures are able to quantify global glaucomatous damage.

In the biostatistical literature, the determination of agreement of repeated measurements on identical measurement scales is discussed controversially.¹,² The use of correlation measures is criticized sharply by the investigators in the second study.² However, if different measurement scales are compared, the use of correlation and regression analyses is characterized as at least of limited use.²,³ To our knowledge no really convincing alternatives have been proposed. Therefore, in

From the ¹Department for Medical Informatics, Biometry, and Epidemiology, Erlangen, Germany; and the ²Department of Ophthalmology, University of Erlangen–Nürnberg, Germany.

Supported by Grant SFB-539 from the Deutsche Forschungsgemeinschaft, Bonn, Germany.

Submitted for publication March 10, 1999; revised July 27, 1999; accepted September 23, 1999.

Commercial relationships policy: N.

Corresponding author: Peter Martus, Institute for Medical Informatics, Biometry and Epidemiology, University of Erlangen–Nürnberg, Waldstr. 6, 91054 Erlangen, Germany. peter.martus@imbe.med.uni-erlangen.de

Copyright © Association for Research in Vision and Ophthalmology

1099
useful in the diagnosis of the glaucomatous diseases. One particular reason for this study was the finding reported in earlier studies, that correlations between the NRRA and sensory measures, which have been observed in patients with moderate or advanced glaucomatous damage, are to a much smaller degree observed among patients in the early state of the disease. This fact may be subject to measurement errors or to the pathogenic mechanisms of the disease. We therefore investigated whether in those patients sensory measures provide any information about glaucomatous damage. We restricted the analysis to measures that were sensitive to global damage, omitting measures exclusively sensitive to localized glaucomatous damage or to the variance of the damage (e.g., perimetric loss variance). We did not include methods that allow differentiation between diffuse and localized loss, such as the pattern deviation probability map.

**METHODS**

**Procedures**

**NRRA.** For all eyes, stereo optic disc photographs were taken using a telecentric fundus camera (Carl Zeiss, Oberkochen, Germany). The diapositives were projected in a scale of 1 to 15. The outlines of the optic disc and optic cup were plotted on paper and morphometrically analyzed. The method has been described in detail previously. The ocular and camera magnifications were corrected according to Littmann’s method, taking into account the anterior corneal curvature and the refractive error. The standard protocol for grading of the optic disc photographs consisted of a check list including the following variables: size and shape of the optic disc, presence of disc hemorrhages, location and extent of alpha and beta zones of peripapillary atrophy, diameter of the retinal arterioles, and visibility of the retinal nerve fiber layer. This protocol was used for evaluation of all optic disc photographs. The photographs of the patients included in the study were mixed with photographs of other patients with glaucomatous or nonglaucomatous optic nerve damage and with photographs of normal subjects. The evaluations were performed in a masked fashion by two examiners who had had experience evaluating optic disc photographs of more than 2000 individuals. The coefficient of variation for the morphometric determination of the optic disc structures had been examined in a previous study. Intraobserver variation coefficients were 0.01 for the assessment of the optic disc area and 0.03 for the measurement of the optic cup area. Interobserver coefficients were 0.03. The neuroretinal rim was calculated as the difference of disc area minus cup area.

**Perimetric Mean Defect.** The perimeter (Octopus 501; Interzeag, Schlieren, Switzerland; 59 measure points, program G1, three phases) was used. Local or diffuse visual field loss was defined according to Bebié et al. (pathologic cumulative perimetric defect curves based on graphical display of ranked local defects compared with the 95th and 99th percentiles of normal curves with identification of localized, diffuse, and broadly distributed visual field losses).

**Flicker Test.** A system with a full-field bowl (58 cm in diameter) and a white flicker light was used. The test was performed under photopic conditions and required no fixation by the subject. The flicker threshold was determined at a constant frequency of 37.1 Hz at a time-average luminance of 10 candelas [cd]/m². The mean luminance of the full-field bowl was corrected by taking into account the pupil diameter and the Stiles-Crawford effect. The contrast sensitivity was assessed using a staircase tracking procedure. The mean value of at least six threshold crossings entered the evaluation.

Both electrophysiological tests were performed with a two-channel Maxwellian view system with a Xenon arc lamp as the light source. The circular field was 32° in diameter in all recordings. With this stimulus system, retinal illuminance is independent of pupil width; therefore, no correction of pupil width was necessary. For peak latency of the blue-on-yellow onset visual evoked potential (VEP), one channel provided a high-contrast, 0.88-cyc/deg square-wave stripe pattern of blue light (460 nm, 3.3 × 10⁻² trolands [td]), the other channel provided a homogeneous yellow adaptation light (570 nm, 1.3 × 10⁻⁴ td) that was superimposed on the stripe pattern. Stimulation was in the onset (200 msec)-offset (500 msec) mode. Recording was monopolar from the inion against the left ear lobe while the right ear lobe was grounded. After amplification (EMP 88 [Electronic Medicine Technique, Pölzl, Munich, Germany], filter: 0.5–70 Hz), 150 sweeps (400 msec in length) were averaged (500-Hz sampling rate). Peak time measurements of the onset responses were made from the moment of pattern onset to the peak of the main negative wave (N1). For amplitude of the black-and-white pattern-reversal electroretinogram (ERG), only one channel of the viewing system was used. The stimulus was a vertical, high-contrast (0.93), black-and-white square-wave stripe pattern with a spatial frequency of 0.88 c/deg. The pattern reversal was square wave and occurred at a frequency of 7.8 Hz. The mean luminance was 4263 photopic td. The responses were recorded with a carbon glide electrode hooked over the subject’s lower eye lid. After amplification (EMP 88, filter: 0.5 Hz–70 Hz, no notch filter) the responses were averaged and stored in a digital computer (IBM-AT, Armonk, NY; sampling rate 1000 Hz, 256 msec sweep, n = 30). Four pattern-reversal responses, and therefore eight amplitudes, were analyzed within one sweep. A subsequent fast Fourier analysis evaluated the amplitude of the second harmonic component of a total of 240 pattern-reversal responses. In both procedures, ERG and VEP, two recordings were made to check for reproducibility.

**Diagnostic Criteria**

The definition of glaucoma was based on the optic disc damage. Criteria were: glaucomatous changes of the optic nerve head such as unusual small NRRA in relation to the optic disc size, an abnormal shape of the neuroretinal rim, cup-to-disc ratios that were higher vertically than horizontally, and/or localized or diffuse retinal nerve fiber layer defects. Visual field loss and intraocular pressure (IOP) were no inclusion criteria. The description of the sample, however, included visual field loss and tonometry. The definition of normal-pressure glaucoma (max IOP, ≤21 mm Hg) and open-angle glaucoma (max IOP, >21 mm Hg) was based on at least two IOP measurements before initial medical therapy.

**Patients and Control Subjects**

**Glaucoma Patients.** Four hundred six eyes of 203 patients with chronic open-angle glaucoma were included in the study (Table 1). The patient group was divided into two subgroups: One subgroup included 109 eyes of 76 patients with an NRRA of at least 1.35 mm² (equivalent to the mean + 1 SD in...
the control group), and the other subgroup contained 297 eyes of 170 patients with NRRA of less than 1.35 mm². One hundred eighty-six of all 203 glaucoma patients had primary open-angle glaucoma, 17 had secondary open-angle glaucoma (9 primary melanin dispersion, 6 pseudoexfoliative syndrome, 1 anterior chamber angle recession after ocular contusion, 1 who developed glaucoma under systemic cortisone therapy). Three hundred fifteen of the 406 glaucomatous eyes were treated topically. On the day of examination, no subject had intraocular pressure more than 24 mm Hg.

**Control Subjects.** For comparison we used a control group of 200 eyes of 100 normal persons from the university staff with intraocular pressure below 21 mm Hg, normal optic discs, and normal visual fields (Table 1). Slit lamp examination and ophthalmoscopy revealed no diseases.

Only patients and control subjects whose eyes were both classified in the same of the two groups were included in the study. All subjects had clear optic media; exclusion criteria were other eye diseases (optic media opacities, retinal diseases) and systemic diseases (diabetes mellitus). Subjects performing visual field testing with false-positive and false-negative responses of more than 12% were excluded (three control subjects, eight patients). All subjects had visual acuity of 20/30 or better. The principles of the Declaration of Helsinki were complied with: The participants were informed about the purpose of the study and the nature of the measurements. They were informed that they had the right to withdraw from the study at any time and signed an informed consent form.

**Statistical Methods**

The normal distribution assumption could be accepted for the NRRA, and the amplitude of the pattern-reversal ERG. For the perimetric mean defect and the flicker sensitivity, the log transformation provided the best adjustment to normality, and for the peak latency of the blue-on-yellow VEP, the log-log transformation provided the best adjustment. The flicker sensitivity, the amplitude of the pattern-reversal ERG, and the peak latency of the blue-on-yellow VEP were age adjusted by linear regression analysis in the control group. Correlation analyses used the Pearson product moment correlation coefficient. If outliers or nonlinearity were present, the Spearman correlation coefficient was computed additionally. In every subject both eyes were used. The necessary corrections for statistical testing and estimation of standard errors were taken into account. For group comparisons of continuous variables the mean value of the left and right eyes were used. All statistical comparisons are versus controls; mean values of left and right eyes were used.

**Table 1. Age, Gender, and Measurements in Control Subjects and Glaucoma Patients**

<table>
<thead>
<tr>
<th></th>
<th>NRRA ≥1.35 mm²</th>
<th>NRRA &lt;1.35 mm²</th>
<th>Full Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (n)</td>
<td>76</td>
<td>170</td>
<td>203</td>
</tr>
<tr>
<td>Eyes (n)</td>
<td>109</td>
<td>297</td>
<td>406</td>
</tr>
<tr>
<td>Gender (M/F)</td>
<td>53/47</td>
<td>49/27</td>
<td>101/102</td>
</tr>
<tr>
<td>Age (y)</td>
<td>44.7 ± 11.5</td>
<td>48.6 ± 10.9§</td>
<td>51.2 ± 11.2§</td>
</tr>
<tr>
<td>Visual field loss (n)</td>
<td>(0/200)</td>
<td>(13/96)</td>
<td>(146/151)</td>
</tr>
<tr>
<td>IOP (mm Hg; n elevated/normal)</td>
<td>17.4 ± 2.1 (0/200)</td>
<td>28.6 ± 9.6§ (22/87)</td>
<td>26.6 ± 7.5§ (93/204)</td>
</tr>
<tr>
<td>VA</td>
<td>1.09 ± 0.12</td>
<td>1.09 ± 0.12</td>
<td>1.06 ± 0.14</td>
</tr>
<tr>
<td>NRRA</td>
<td>1.65 ± 0.3 (1.43, 1.82)</td>
<td>1.61 ± 0.26 (1.43, 1.71)</td>
<td>0.96 ± 0.27§ (0.78, 1.18)</td>
</tr>
<tr>
<td>MD</td>
<td>1.05 ± 1.25 (0.20, 1.79)</td>
<td>1.75 ± 3.03 (0.18, 2.35)</td>
<td>4.13 ± 4.87§ (0.84, 5.74)</td>
</tr>
<tr>
<td>FLI</td>
<td>1.46 ± 0.17 (1.35, 1.59)</td>
<td>1.37 ± 0.21 (1.25, 1.51)</td>
<td>1.29 ± 0.26 (1.16, 1.46)</td>
</tr>
<tr>
<td>VEP</td>
<td>117.5 ± 8.4 (112.8, 122.5)</td>
<td>120.9 ± 10.6 (113.8, 125.8)</td>
<td>124.0 ± 14.0§ (115.3, 134.5)</td>
</tr>
<tr>
<td>ERG</td>
<td>3.78 ± 1.08 (3.01, 4.48)</td>
<td>3.62 ± 1.01 (3.01, 4.12)</td>
<td>3.09 ± 0.98 (2.35, 3.69)</td>
</tr>
</tbody>
</table>

Values are expressed as mean ± SD of raw values (without transformations, see the Statistical Methods section). For NRRA, MD, FLI, ERG, and VEP, upper and lower quartiles are given in parentheses. VA, visual acuity; NRRA, neuroretinal rim area; MD, perimetric mean defect; FLI, flicker test, log temporal contrast sensitivity; ERG, amplitude of the pattern-reversal electroretinogram; VEP, peak latency of the blue-on-yellow visually evoked potential.

* Forty-three patients had one eye with NRRA of less than 1.35 mm² and one eye with NRRA of at least 1.35 mm². All statistical comparisons are versus controls; mean values of left and right eyes were used.

† P ≤ 0.05.
‡ P ≤ 0.01.
§ P ≤ 0.001.
The Method of Confirmatory Factor Analysis

Confirmatory factor analysis is a special case of structural equation modeling. This method is a standard statistical approach in other fields of applications, especially in the psychometric literature. However, because to our knowledge with only a few exceptions these or related approaches are rarely applied in ophthalmology, we provide a short description of how the method works.

Assuming that a gold standard exists that perfectly measures glaucomatous damage, then any diagnostic procedure could be judged by the size of the correlation \( r \) between this procedure and the perfect gold standard. If the value of this correlation lies near 1, the procedure would measure glaucomatous damage very accurately. Now, assume that we knew for two different measures the correlations with this gold standard, say \( r_1 \) and \( r_2 \). Then it can be shown by simple calculations that the correlation between the two measures, say \( r_{12} \), would be at least equal to the product \( r_1 \cdot r_2 \). For example, if the first measure correlated to the gold standard with \( r_1 \) equal to 0.8 and the second measure with \( r_2 \) equal to 0.9, then the correlation \( r_{12} \) between both measures would be at least 0.72. However, under certain circumstances a correlation \( r_{12} \) would be considerably larger than \( r_1 \cdot r_2 \). This would be the case, if the two measures were also influenced by common factors different from the glaucoma disease. If both measures were psychophysical, such a common factor could be, for example, the concentration of the patient. The dependency of both measures on the concentration of the patient would increase the correlation coefficient. This, of course, would not increase the validity of both measures concerning the underlying disease. Procedures that do not share common factors different from glaucomatous damage and for which therefore \( r_{12} \) equals exactly \( r_1 \cdot r_2 \), are called conditionally independent.

If a perfect gold standard does not exist, only the pairwise correlations \( r_{12} \) between diagnostic measures can be determined. However, if at least three measures are conditionally independent, then it is possible to calculate from the pairwise correlations \( r_{12} \), \( r_{13} \), \( r_{23} \), the values of \( r_1 \), \( r_2 \), and \( r_3 \) in absence of any gold standard procedure. The system of equations

\[
\begin{align*}
    r_{12} &= r_1 \cdot r_2 \\
    r_{13} &= r_1 \cdot r_3 \\
    r_{23} &= r_2 \cdot r_3 \\
\end{align*}
\]

is exactly solved by

\[
\begin{align*}
    r_1 &= \sqrt{[r_{12} \cdot r_{13}] / r_{23}} \\
    r_2 &= \sqrt{[r_{12} \cdot r_{23}] / r_{13}} \\
    r_3 &= \sqrt{[r_{13} \cdot r_{23}] / r_{12}}.
\end{align*}
\]

The solutions \( r_1 \), \( r_2 \), and \( r_3 \) are usually called factor loadings or path coefficients. These path coefficients have values between 0 and 1. A path coefficient of 0 shows that the measurement does not contain any information about the underlying disease. A path coefficient of 1 corresponds to a measurement that perfectly quantifies the disease.

However, the crucial assumption of conditional independence cannot be tested for only three measures. Only for more than three measures can the suitability of the model be examined. With, for example, five procedures, 10 correlations are available, and the system (1) contains 10 equations for only five unknowns. Moreover, with more than three measures the crucial assumption may be weakened to a certain degree. There may be some additional correlations quantified between the measures because of factors distinct from the glaucoma disease. The price is that exact formulae are not available, and numerical algorithms have to be used. In addition to the determination of the path coefficient, the model also allows the computation of an index, quantifying the damage for the individual patient. This index comprises the best approximation of a gold standard by using the procedures under investigation.

In summary, the size of the path coefficients obtained in our analysis shows the ability of the measurements to quantify glaucomatous damage. They preserve a ranking of these measurements. The measurement that shows the largest coefficient is able to quantify glaucomatous damage best. The conditional independence assumption is essential in our approach.

Confirmatory Analysis for Global Glaucomatous Damage

The basis of the analysis is the fact that all five procedures under investigation measure global glaucomatous damage by using clearly different approaches. With the exceptions listed below, we therefore assume that the observed correlations are due to the global glaucomatous damage. However, three limitations have to be respected: 1) The decrease of the NRRA may be observable before sensory damages, 2) perimetry and the flicker test may also correlate among each other because of some general psychophysical fitness, and 3) the electrophysiological procedures (pattern-reversal ERG, blue-on-yellow VEP) may also correlate, because the same technical device is used (Fig. 1). The assumptions of conditional independence were checked in several ways: First, we performed correlation analyses in the control group. In this group, no glaucomatous damage is present. Therefore, we do not expect correlations between diagnostic measures of glaucoma within this group. In contrast, if we observe such correlations, factors different from glaucoma influence the respective diagnostic measures. A similar analysis concerning only the correlations of the rim area with functional test results has been performed previously. Second, for all pairs of measures we compared the correlations in the sample to the correlations predicted by the model. This answers the question of whether the smaller set of path coefficients is able to explain the larger set of pairwise correlations. Third, for all possible groups of three and four variables we computed the path coefficients and compared the different results. If the path coefficients differed in these analyses, this would contradict the assumptions of our model.

The global goodness of fit of the models was examined using the \( \chi^2 \) test.\[^{14} \]

RESULTS

Description and Comparisons between Groups

Descriptive analyses are given in Table 1. Box plots of the procedures are given in Figure 2A through 2E. There was no difference in gender (\( P > 0.6 \)) between patients and control subjects. Glaucoma patients were 6.5 years older than the control subjects (\( P < 0.001 \)). The visual acuity in the control group and the glaucoma groups was not significantly different (\( P > 0.1 \)). All diagnostic measurements were significantly different in the group with reduced NRRA compared with the control group (\( P < 0.001 \)). Significant differences between the control subjects and the subgroup with NRRA of at least 1.35
NRRA, the perimetric mean defect (not for peak latency of the blue-on-yellow VEP) with a range of correlations within the whole patient group were significant activity and perimetric mean defect (only between the two psychophysical measures, flicker sensitivity). In the control group significant correlations were observed for the correlation between the perimetric mean defect and the flicker sensitivity. There was a significant additional correlation between both these measures but not between the two electrophysiological measures, peak latency of the blue-on-yellow VEP and amplitude of the pattern-reversal ERG. Inspection of the path coefficients showed a ranking of the measurements with the perimetric mean defect clearly representing best the damage of glaucomatous disease. The flicker test and the area of the neuroretinal rim followed, whereas the pattern-reversal ERG amplitude and blue-on-yellow VEP peak latency showed the lowest correlation to the common factor, global glaucomatous damage.

All pairwise correlations between the five measures were predicted by the model with a deviation below 0.05. The greatest difference appeared for the correlation between the flicker sensitivity and the pattern-reversal ERG amplitude. The observed correlation was 0.355, and the predicted correlation was 0.313. The investigation of 12 different subgroups of the five measurements (Table 4) demonstrated only moderate scatter of the path coefficients, which supported the validity of the model. The index of global damage obtained from the analysis was (0.49 · MD) + (0.06 · FLI) + (0.24 · NRRA) + (0.17 · ERG) + (0.03 · VEP) for standardized variables, where MD is perimetric mean defect and FLI is flicker test, log temporal contrast sensitivity.

Confirmatory Factor Analysis II

Subgroup with NRRA of Less Than 1.35 mm². This additional analysis was performed to examine how much our results may have depended on the stage of glaucomatous disease present in the sample. The appropriateness of the analysis seemed to be even higher than in analysis I (goodness of fit: $P = 0.34$), but this was mainly due to the reduced sample size in analysis II compared with that in analysis I. When the sample size independent Holter criterion is used for reference (results not given), both models showed nearly identical fit. Again there was a significant additional correlation between perimetric mean defect and flicker sensitivity but not between the peak latency of the blue-on-yellow VEP and the amplitude of the pattern-reversal ERG. The path coefficients of the NRRA, the perimetric mean defect, and the flicker sensitivity were higher in analysis II than in analysis I, whereas for the electrophysiological procedures there were no differences between both models (Table 3). All pairwise correlations between the five measures were predicted by the model with a deviation below 0.04 (Table 2). The greatest difference again appeared for the correlation between the flicker sensitivity and the
FIGURE 2. (A through E) Box plots of diagnostic measures in the control subjects, the full sample of patients, and the subsamples with NRRA of less than or at least 1.35 mm². (A) NRRA, neuroretinal rim area; (B) MD, perimetric mean defect; (C) FLI, flicker test, log temporal contrast sensitivity; (D) ERG, amplitude of pattern-reversal electroretinogram; and (E) VEP, peak latency of blue-on-yellow VEP. The central line in the middle of the rectangle indicates the median, the top and bottom of the box indicate the 25% and 75% quartiles, respectively. The antennae are defined as those observations larger than the 25% quartile $\times 1.5$ box length and those smaller than the 75% quartile $+ 1.5 \times$ box length. For normally distributed data, approximately 99.3% of the observations lie in the range defined by the upper and lower antenna. The circles indicate extreme observations. Asterisks indicate outliers.
amplitude of the pattern-reversal ERG: The observed correlation was 0.392, and the predicted correlation was 0.356. The scatter of coefficients in 12 different subgroups of the five measurements (Table 4) was nearly identical in analysis I and analysis II. The index of global damage obtained from the reduced sample was $(0.53 \cdot \text{MD}) + (0.07 \cdot \text{FLI}) + (0.26 \cdot \text{NRRA}) + (0.13 \cdot \text{ERG}) + (0.15 \cdot \text{VEP})$.

**Subgroup with NRRA of at Least 1.35 mm$^2$.** In this group a path analysis was not adequate, as can be seen from Table 2 and also because of the small size of the sample.
Therefore, the structure depicted in Figure 1 may not be true. However, correlation analysis showed a strong association between the flicker sensitivity and the perimetric mean defect and to some lower degree between the perimetric mean defect and the peak latency of the blue-on-yellow VEP (Table 2, Fig. 5). The correlation between the flicker sensitivity and the perimetric mean defect was substantially higher than that expected from the analyses in the control group. Therefore, to some degree, even in this subsample the measurements provided information about the early stage of glaucomatous damage.

**TABLE 2.** Pairwise Correlations in Control Subjects and Glaucoma Patients: Observed Correlations and Values Predicted by Confirmatory Factor Analysis

<table>
<thead>
<tr>
<th>Procedures</th>
<th>Control Subjects</th>
<th>NRRA ≥ 1.35 mm²</th>
<th>NRRA &lt; 1.35 mm²</th>
<th>Full Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects (n)</td>
<td>100</td>
<td>76*</td>
<td>170*</td>
<td>203*</td>
</tr>
<tr>
<td>Eyes (n)</td>
<td>200</td>
<td>109</td>
<td>297</td>
<td>406</td>
</tr>
<tr>
<td>Method</td>
<td>Pearson</td>
<td>Pearson</td>
<td>Pearson</td>
<td>Spearman</td>
</tr>
<tr>
<td>NRRA–MD</td>
<td>-0.04</td>
<td>0.00</td>
<td>0.64‡ (0.64)</td>
<td>0.53‡ (0.52)</td>
</tr>
<tr>
<td>NRRA–FLI</td>
<td>0.00</td>
<td>0.05</td>
<td>0.44‡ (0.48)</td>
<td>0.34‡ (0.38)</td>
</tr>
<tr>
<td>NRRA–VEP</td>
<td>-0.03</td>
<td>-0.08</td>
<td>0.40‡ (0.40)</td>
<td>0.33‡ (0.35)</td>
</tr>
<tr>
<td>NRRA–ERG</td>
<td>-0.05</td>
<td>-0.02</td>
<td>0.39‡ (0.40)</td>
<td>0.36‡ (0.34)</td>
</tr>
<tr>
<td>MD–FLI</td>
<td>0.28‡</td>
<td>0.53‡</td>
<td>0.68‡ (0.68)</td>
<td>0.66‡ (0.66)</td>
</tr>
<tr>
<td>MD–VEP</td>
<td>0.15</td>
<td>0.27</td>
<td>0.46‡ (0.47)</td>
<td>0.45‡ (0.45)</td>
</tr>
<tr>
<td>MD–ERG</td>
<td>0.07</td>
<td>0.05</td>
<td>0.47‡ (0.48)</td>
<td>0.42‡ (0.43)</td>
</tr>
<tr>
<td>FLI–VEP</td>
<td>0.00</td>
<td>0.06</td>
<td>0.38‡ (0.35)</td>
<td>0.33‡ (0.32)</td>
</tr>
<tr>
<td>FLI–ERG</td>
<td>0.08</td>
<td>0.16</td>
<td>0.39‡ (0.56)</td>
<td>0.35‡ (0.31)</td>
</tr>
<tr>
<td>VEP–ERG</td>
<td>0.12</td>
<td>0.17</td>
<td>0.31‡ (0.50)</td>
<td>0.31‡ (0.29)</td>
</tr>
</tbody>
</table>

Pearson product moment correlations. Parentheses: Correlations predicted by confirmatory factor analysis (see Table 3). For abbreviations see Table 1. Log-transformation for MD and FLI; log–log transformation for VEP. The variables are rescaled so that higher values show presence of disease.

* Forty-three patients had one eye with NRRA of less than 1.35 mm² and one eye with NRRA of at least 1.35 mm².
† $P \leq 0.01$.
‡ $P \leq 0.001$. 

---

**Figure 2.** (Continued)
DISCUSSION

In the present study, we examined four sensory parameters and one morphologic procedure that have been shown in earlier studies to measure global glaucomatous damage. However, it should be noted that some investigators have found that in many cases localized visual field loss would be an early indicator of beginning glaucoma. Whereas in other cross-sectional studies some procedures (e.g., perimetry and morphometry) are preselected as a gold standards, in the present analysis established methods and new experimental procedures were given identical chances to prove validity. Therefore, perimetrically detectable visual field loss was not an inclusion criterion in our sample. However, the grouping of our sample in patients and control subjects relied on the gold standard morphologic damage. The definition of glaucoma itself has been discussed widely. In the Rotterdam Eye Study, two different definitions were used to account for this problem. Therefore, in our analysis we also investigated to what degree a bias might have resulted from this design.

There have been only a few applications of factor analysis or related statistical methods in ophthalmology. In one study, a model with a priori hypotheses was performed to explain multistage mechanisms of activities of daily living after cataract surgery. To our knowledge in all other applications, explor-
Table 3. Results of Confirmatory Factor Analysis

<table>
<thead>
<tr>
<th></th>
<th>NRRA</th>
<th>MD</th>
<th>FLI</th>
<th>VEP</th>
<th>ERG</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Path coefficient*</td>
<td>0.64 ± 0.07</td>
<td>0.81 ± 0.08</td>
<td>0.59 ± 0.07</td>
<td>0.55 ± 0.07</td>
<td>0.54 ± 0.06</td>
</tr>
<tr>
<td><strong>Subsample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Path coefficient*</td>
<td>0.73 ± 0.05</td>
<td>0.87 ± 0.05</td>
<td>0.65 ± 0.08</td>
<td>0.54 ± 0.07</td>
<td>0.55 ± 0.07</td>
</tr>
</tbody>
</table>

For abbreviations see Table 1. The full sample included 406 eyes of 203 patients with open-angle glaucoma. The subsample included 297 eyes of 170 glaucoma patients with NRRA less than 1.35 mm². Log transformation for MD and FLI; log-log transformation for VEP. The variables are rescaled so that higher values show presence of disease.

* The path coefficients (±SE) measure the validity of each procedure in quantifying global glaucomatous damage. A value of 0 corresponds to a measure which does not contain any information about the underlying disease, a value of 1 corresponds to a measure which perfectly quantifies the disease. SE determined by the bootstrap method (3000 replications), correction factor to a measure, which does not contain any information about the underlying disease, a value of 1 corresponds to a measure that perfectly quantifies the disease. The rows correspond to twelve separate confirmatory factor analyses in twelve different subgroups of the five measurements. The path coefficients measure the validity of each procedure in quantifying global glaucomatous damage. A value of 0 corresponds to a measure that contains no information about the underlying disease; a value of 1 corresponds to a measure that perfectly quantifies the disease. The best result was obtained for the perimetric mean defect, followed by the area of the neuroretinal rim and flicker sensitivity. The two electrophysiological measures, amplitude of the pattern-reversal ERG and peak latency of the blue-on-

Table 4. Sensitivity Analysis for Different Subgroups of Measurements

<table>
<thead>
<tr>
<th>Path Coefficients of Selected Measurements</th>
<th>NRRA</th>
<th>MD</th>
<th>FLI</th>
<th>VEP</th>
<th>ERG</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRRA, MD, VEP</td>
<td>0.62</td>
<td>0.85</td>
<td>—</td>
<td>0.53</td>
<td>—</td>
</tr>
<tr>
<td>NRRA, MD, ERG</td>
<td>0.67</td>
<td>0.79</td>
<td>—</td>
<td>—</td>
<td>0.53</td>
</tr>
<tr>
<td>NRRA, FLI, VEP</td>
<td>0.58</td>
<td>—</td>
<td>0.58</td>
<td>0.57</td>
<td>—</td>
</tr>
<tr>
<td>NRRA, FLI, ERG</td>
<td>0.58</td>
<td>0.58</td>
<td>—</td>
<td>—</td>
<td>0.61</td>
</tr>
<tr>
<td>NRRA, VEP, ERG</td>
<td>0.62</td>
<td>—</td>
<td>—</td>
<td>0.54</td>
<td>0.58</td>
</tr>
<tr>
<td>MD, VEP, ERG</td>
<td>—</td>
<td>0.78</td>
<td>—</td>
<td>0.58</td>
<td>0.54</td>
</tr>
<tr>
<td>FLI, VEP, ERG</td>
<td>—</td>
<td>—</td>
<td>0.62</td>
<td>0.54</td>
<td>0.57</td>
</tr>
<tr>
<td>NRRA, MD, FLI, VEP*</td>
<td>0.62</td>
<td>0.85</td>
<td>0.58</td>
<td>0.53</td>
<td>—</td>
</tr>
<tr>
<td>NRRA, MD, FLI, ERG*</td>
<td>0.66</td>
<td>0.79</td>
<td>0.56</td>
<td>—</td>
<td>0.54</td>
</tr>
<tr>
<td>NRRA, MD, VEP, ERG</td>
<td>0.65</td>
<td>0.81</td>
<td>—</td>
<td>0.55</td>
<td>0.53</td>
</tr>
<tr>
<td>NRRA, FLI, VEP, ERG</td>
<td>0.59</td>
<td>0.59</td>
<td>0.59</td>
<td>0.54</td>
<td>0.53</td>
</tr>
<tr>
<td>MD, FLI, VEP, ERG*</td>
<td>0.78</td>
<td>0.62</td>
<td>—</td>
<td>0.57</td>
<td>0.54</td>
</tr>
<tr>
<td><strong>Subsample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range of path coefficients</td>
<td>0.59-0.67</td>
<td>0.78-0.85</td>
<td>0.56-0.62</td>
<td>0.53-0.57</td>
<td>0.53-0.61</td>
</tr>
<tr>
<td><strong>Subsample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range of path coefficients</td>
<td>0.66-0.75</td>
<td>0.83-0.89</td>
<td>0.62-0.69</td>
<td>0.53-0.59</td>
<td>0.54-0.59</td>
</tr>
</tbody>
</table>

For abbreviations see Table 1. Full sample included 406 eyes of 203 patients with chronic open-angle glaucoma. Subsample included 297 eyes of 170 glaucoma patients with NRRA less than 1.35 mm². Log transformation for MD and FLI; log-log transformation for VEP. The variables are rescaled so that higher values show presence of disease. The path coefficients measure the validity of each procedure in quantifying global glaucomatous damage. A value of 0 corresponds to a measure that contains no information about the underlying disease; a value of 1 corresponds to a measure that perfectly quantifies the disease. The rows correspond to twelve separate confirmatory factor analyses in twelve different subgroups of the five measurements. The path coefficients measure the validity of each procedure in quantifying global glaucomatous damage.
yellow VEP, showed the lowest, nearly identical coefficients. Additionally, in the present analysis we considered the possibility of common factors different from the glaucoma disease that might increase the correlations between the procedures under investigation. The analyses clearly showed that there was indeed a common factor shared by flicker testing and perimetry that we termed psychophysical fitness. The same was not true, however, for the two electrophysiological procedures, pattern-reversal ERG and blue-on-yellow VEP, although the same technical device was used.

The splitting of the patient group into two subsamples was performed by using a cutoff point of 1 SD from the mean value of the control group. This cutoff point is probably not the same as would be chosen in a study on diagnostic discrimination of glaucoma patients from control subjects with no disease. Inspection of Figure 4 shows, however, that this choice is useful, because the statistical assumptions of correlation analyses are clearly better fulfilled in the subsample compared with the complete group. However, the ranking of procedures according to their path coefficients was identical in both analyses.

There was a significant age difference of 6 years between the control subjects and the glaucoma patients. However, all diagnostic measures were age corrected by linear regression analysis in the control group. Dichotomizing age at the median of 47 years did not lead to any significant differences between younger (at most, 46 years) and elder (at least, 47 years) control probands (results not given in detail). Furthermore, the control group served only to explore unexpected correlations between the diagnostic procedures. It was not the intent of this study to prove how well the procedures were able to discriminate between patients and healthy control subjects. However, we cannot completely exclude the possibility that the concentration and compliance of control subjects who were recruited from the university staff was higher than that of the patients.

The classification of the subjects was according to qualitative morphologic criteria. These criteria are based on the same photograph as the measurement of the NRRA, which could have produced a statistical bias. However, parameter estimates did not differ systematically in the models with NRRA compared with those without (Tables 4A, 4B, rows 6, 7, 12). Therefore, we did not expect a bias concerning the morphologic variable rim area, because of the morphologic selection criteria.

There was a clear discrepancy between the rather high path coefficients of the flicker sensitivity and the low weight of this measure in the final index for both models. The reason for this lies in the high correlation between the flicker sensitivity and the perimetric mean defect. If the perimetric mean defect is excluded from the analysis, the path coefficients do not change considerably (Tables 4A, 4B) but now in the index the flicker test is assigned the highest weight among the remaining four procedures (results not given). Therefore, it is the path coefficient, not the weight in the index that truly reflects the usefulness of a procedure.

The perimetric mean defect was the best variable in quantifying global glaucomatous damage. However, in the present study, indeed only less than half of all glaucomatous eyes showed localized or diffuse visual field loss. For the perimetric mean defect there was no significant difference between control subjects and the subgroup of glaucoma patients with a rim area of at least 1.35 mm² (Table 1). This result is in accordance with the literature. A growing number of histologic and clinical studies have convincingly shown that optic nerve damage in patients with glaucoma occurs and can be detected before conventional perimeter uncovers early visual field defects. Clinical investigations using morphologic techniques

**FIGURE 5.** Regression analysis of sensory procedures in the subsample with NRRA of at least 1.35 mm². For abbreviations see Figure 1. Linear regression analysis of FLI, VEP, and ERG on MD. Correlations (Spearman) between MD and FLI and between MD and VEP were significant. All procedures were rescaled so that higher values indicate pathologic results. FLI, ERG, and VEP were standardized relative to the mean and SD of the control group. The vertical line marks an MD (raw value) of 2.3 dB which is equivalent to mean +1 SD in the control group, the horizontal line marks mean +1 SD in the control group for the three measures FLI, ERG, and VEP.
have shown that quite a number of optic nerve head variables, such as the neuroretinal rim as a whole and measured separately in various disc sectors, the shape of the neuroretinal rim, and the presence and size of peripapillary atrophy, were abnormal in some individuals with ocular hypertension but normal findings in conventional visual field examinations.\(^2\) In contrast, the flicker test showed a significant difference between the control group and the group with NRRA at least 1.35 mm\(^2\) but was inferior to perimetry in quantifying global glaucomatous damage. Both results are not contradictory.

The reason that the perimetric mean defect behaves very well in the quantification of glaucomatous damage but much poorer in the early diagnosis of the disease lies in the great overlap of the pathologic and the normal range (4 dB) in perimetry. This has much less consequence for correlation analyses within patients than for early diagnosis of glaucoma. In the subsample with NRRA of at least 1.35 mm\(^2\), we observed no correlation between the NRRA and sensory measures. Nevertheless, the analyses showed that within this subsample compared with the healthy control group the perimetric mean defect and the flicker sensitivity correlated with each other to a much higher degree, but not to the rim area. One explanation is that morphologic diagnosis of glaucoma is not based on the NRRA alone, but also on other morphologic criteria, such as the presence of localized retinal nerve fiber layer defects. We therefore conclude that also in this study group with rim area at least 1.35 mm\(^2\) the psychophysical measures quantify global glaucomatous damage to a certain degree. Measures that reflect the status of disease to only a moderate degree may still be useful in early diagnosis. Also the amplitude of the ERG and the peak latency of the VEP did not show significant differences between the control subjects and the patients with NRRA of at least 1.35 mm\(^2\). Therefore, at least in patients with “normal” NRRA (but other morphologic signs that indicate glaucoma) both measures are not sensitive in early diagnosis of glaucoma.

One purpose of this study was to demonstrate the usefulness of confirmatory factor analysis for the problem of validating measures of glaucomatous damage. The method is most powerful if a whole group of measurements is available but a true gold standard does not exist. We believe that this situation is exactly met by many studies concerning primary open angle glaucoma. In their most simple form, formulae (2) allow evaluation of any electrophysiological procedure if quantitative morphologic data on optic nerve damage and results of a psychophysical test of global glaucomatous damage are available. This approach improves the mere calculation of correlations substantially.

References